today: continue MCMC
        Gibbs sampling & M.H.

continue Markov chain results

def: irreducible Markov chain \iff there exists a positive prob "path"
      \text{from every state}
      \forall (i, j), \exists an integer m_{ij} \text{ st. } (A^m)_{ij} > 0

(by Perron-Frobenius thm.) \Rightarrow irreducible M.C. has a unique stationary dist.

* in order to converge to it, we need aperiodicity as well

irreducible and aperiodic M.C. \iff \exists an integer m \text{ s.t. } A^m > 0

aka. regular M.C.,
or ergodic M.C.

(finite state)

\( \exists m \text{ s.t. } A^m > 0 \)

\( \Rightarrow \) [note: a sufficient condition for an irreducible M.C. to be aperiodic is \( \exists i \) s.t. \( A_i > 0 \)]

\[
A = \begin{pmatrix}
0 & 1/2 & 1/2 \\
1/2 & 0 & 1/2 \\
1/2 & 1/2 & 0
\end{pmatrix} = \frac{1}{2} (I - W^T)
\]
Thm. If a finite M.C. is ergodic (regular),

then it has a unique stationary dist. \( \pi \)

and for any starting dist. \( \pi_0 \), \( \lim_{t \to \infty} A^t \pi_0 = \pi \)

the speed of convergence is related to the mixing time \( \tau \) of the chain

\[ \tau \triangleq \frac{1}{1 - \lambda_2(A)} \]

where \( \lambda_2(A) \) is the second biggest eigenvalue of \( A \)

\[ ||A^t \pi_0 - \pi||_1 \leq C \exp(-t/\tau) \]

* Intuition (from linear algebra) [informal argument]

Simpler case, suppose \( A \) is diagonalizable with orthogonal matrix \( U \)

\( A = U \Sigma U^T \) with \( \Sigma = \begin{pmatrix} \lambda & 0 \\ 0 & 0 \end{pmatrix} \)

\( U \rightarrow \) basis of evecetors (by von Neumann-Frobenius thm.)
\[ U = [u_1, \ldots, u_k] \]  
\[ \text{take } U = \pi / \| \pi \|_2 \]  
\[ \begin{bmatrix} \lambda_1 \iff \lambda_2 \iff \ldots \iff \lambda_k \end{bmatrix} \]  
\[ \text{can show that } \lambda_1 > |\lambda_2| > \ldots > |\lambda_k| \]

Let \( \alpha_0 \) be coordinate in basis \( U \) of \( \pi \). \[ U U^T = I = U U^T \]
\[ \pi_0 = U \alpha_0 \]  
(i.e. \( \alpha_0 = U^T \pi_0 \)  
\[ (\alpha_0)_1 = \frac{\pi_0^T \pi_0}{\| \pi_0 \|^2} \])

\[ A^T \pi_0 = (U \Sigma U^T)(U \alpha_0) \]
\[ = U \Sigma \alpha_0 \]
\[ = U \alpha_0 \]
\[ \Sigma^T = \begin{pmatrix} \lambda_1^T & \lambda_2^T & \ldots & \lambda_k^T \end{pmatrix} \]

\[ A^T \pi_0 = \begin{pmatrix} \alpha_0 & \alpha_0 & \ldots & \alpha_0 \end{pmatrix} \begin{pmatrix} \lambda_1 \alpha_0 \lambda_2 \alpha_0 \ldots \lambda_k \alpha_0 \end{pmatrix} \]
\[ \rightarrow 0 \]

because \( A^T \pi_0 = 1 \) \( \forall \pi \Rightarrow (\alpha_0)_1 = 1 \)  
[not fully correct]

\[ \| A^T \pi_0 - \pi_0 \|_2 \leq \lambda_1 \| \pi_0 \|_2 \]  
\[ |\lambda_1| = 1 - \varepsilon_1 \]
\[ \varepsilon_1 \leq 1 - |\lambda_1| \]
\[ |\lambda_1| = \exp(-\varepsilon_1) \]
\[ 1 - x \leq \exp(-x) \]  
\[ \forall x \]
\[ |\lambda_1| \leq \exp(-\varepsilon_1) \]
\[ \Rightarrow \frac{\varepsilon_1}{\lambda_1} \]

- mixing time is often exponentially big

---

\( \varepsilon \) How do we design \( A \) so \( A^T \pi_0 \rightarrow \pi_0 \)?
one “easy way”

reversible M.C. \[ \iff \exists \text{ dist. } \pi \text{ s.t. } A_{ij} \pi_j = A_{ji} \pi_i \quad \forall (i,j) \]

“detailed balance equation”

\[
\text{it means that } P^T x_t = i, \quad x_{t-1} = j \implies P^T x_t = j, \quad x_{t-1} = i \quad \frac{\pi_j}{\pi_i}.
\]

sufficient condition (but not necessary)

for \( \pi = \pi \) 

proof: \( (A \pi)_i = \sum_j A_{ij} \pi_j = \sum_j A_{ji} \pi_i = \pi_i (\sum_j A_{ij}) \)
[i.e. if in state $x$, we sample $x' \mid x \sim q(x' \mid x)$]

\[
\alpha(x' \mid x) \triangleq \min \left\{ 1, \frac{q(x' \mid x) p(x')}{q(x \mid x') p(x)} \right\}
\]

accept new state $x'$ with prob.

if reject \implies stay in same state $x$

[This still a new sample]

vs rejection sampling

where only "accepted states" are sample

acceptance ratio to satisfy detailed balance

\[
\begin{aligned}
\text{MH alg.:} \\
\text{start at } x(0) \\
\text{for } t = 1, \ldots \\
\text{propose } x^{(t)} \sim q(x' \mid x^{(t-1)}) \\
\text{flip a biased coin with prob } \alpha(x^{(t)} \mid x^{(t-1)}) \text{ to be 1} \\
\text{if accept (coin = 1):} \\
\quad \text{let } x^{(t)} = x^{(t)} \\
\text{else: } x^{(t)} = x^{(t-1)}
\end{aligned}
\]

end for

note: for symmetric $q(x' \mid x)$, always accept if $p(x') \geq p(x)$

\[\text{[Metropolis alg.]} \rightarrow \text{like a noisy hill-climbing alg.}\]
[Metropolis alg.]

I verify as exercise that it satisfies detailed balance.

* For convergence: if M.H. chain is aperiodic, then we converge to correct unique stationary dist. \( \pi \)

Sufficient conditions:

\( q(\cdot|x) > 0 \) for irreducibility

\( q(x|\cdot) > 0 \) for aperiodicity

Aside: it is still ok to change proposal with time (inhomogeneous M.C.) \( q_t(x'|x) \)

As long as choice of \( q_t \) does not depend on \( x \) \((t=1)\)

Then convergence theory above will go through (i.e. detailed balance, etc. will give right stationary dist.)

Example:

Suppose \( p \) is a multivariate normal

\( \frac{1}{p} \quad q(x'|x) = N(x'|x, \sigma^2 I) \)
Gibbs sampling alg.

\[ \Rightarrow \text{M.H. with a clear choice of proposal } q_t(x'|x) \]

- **Examples of applications**:
  - **UGM**: \[ \hat{p}(x) = \frac{1}{Z} \varphi(x) \]
  - **Difficult conditional in DGM**
    \[ \hat{p}(x) = \frac{p(x, \bar{x}_E) s(x, \bar{x}_E)}{p(x, \bar{x}_E)} \propto p(x | \bar{x}_E) \]

UGM:

**Cyclic Gibbs sampling alg.**:
- Nodes \( i = 1, \ldots, n \)
- Start at some \( \bar{x}^{(0)} \)
- \[ \text{repeat } l = 1 \]
Convergence of G-S:

\[ a(x(t), x(t+1)) = \frac{q_t(x(t), x(t+1))}{p_t(x(t), x(t+1))} \]

Acceptance ratio:

G-S is M-H with a time-varying proposal.

Then proposed is \( a_t(x(t), x(t+1)) = \frac{p(x_t, x_{t+1})}{q(x_t, x_{t+1})} \).

Suppose we pick \( i \) at time \( t \).

Sample \( x_i = x_i(t+1) \).

Set \( x_i(t) \) for \( j \neq i \).

Pick \( i \) such that

\[ x_i(t+1) \equiv \frac{1}{i} \mod \Omega \]

Then conditional as

For rest to be constant.
Convergence of G.S.:

Let \( A \) be Markov transition kernel of one full cycle of G.S. (i.e. \( n \) steps)

\[ p(x, y) \rightarrow \text{homog. M.C.} \]

\[ A_{ij} > 0 \quad \forall i, j \]

(since can get to any state with \( n \) flips)

\[ \text{supposing } p(x_i | x_{-i}) > 0 \quad \forall x_i \neq x_{-i} \]

\[ \Rightarrow A_t \rightarrow p \]

\[ \text{also works for random scan} \quad \text{pick } i \text{ in } \text{Unif}(1:n) \text{ at each step} \]

**Example:** G.S. for Ising model

**Ising model** \( x_i \in \{0, 1\} \)

**UGM:**

\[ p(x) = \frac{1}{Z(n)} \exp \left( \sum_i m_i x_i + \sum_{i,j \in E} m_{ij} x_i x_j \right) \]

for G.S.,

want to compute \( p(x_i | x_{-i}) \) or \( p(x_i, x_{-i}) \)

\[ = \exp \left( m_i x_i + \sum_{j \in E} m_{ij} x_i x_j + \text{const} \right) \]
\[ p(x_i=1 | x_{-i}) = \frac{\exp\left(\eta_i + \sum_{j \in \text{env}(i)} m_{ij} x_j\right) \cdot \exp(a_{ix})}{1 + \exp\left(11\right)} \cdot \exp(a_{ix}) \]

\[ = \sigma\left(\eta_i + \sum_{j \in \text{env}(i)} m_{ij} x_j\right) \]

⇒ renormalize to get conditional